53745
44

Evaluation of the 29-km Eta Model. Part I: Objective Verification at Three Selected Stations.

Paul A. Nutter[1,2] and John Manobianco[1,2]

[1]ENSCO, Inc.
Cocoa Beach, FL

[2]Applied Meteorology Unit
NASA / Kennedy Space Center, FL

Submitted to:
Weather and Forecasting
March 1998

Corresponding author address: John Manobianco, ENSCO, Inc., 1980 N. Atlantic Ave., Suite 230, Cocoa Beach, FL 32931, E-mail: johnm@fl.ensco.com.

Abstract

This paper describes an objective verification of the National Centers for Environmental Prediction (NCEP) 29-km eta model from May 1996 through January 1998. The evaluation was designed to assess the model's surface and upper-air point forecast accuracy at three selected locations during separate warm (May - August) and cool (October - January) season periods. In order to enhance sample sizes available for statistical calculations, the objective verification includes two consecutive warm and cool season periods.

Systematic model deficiencies comprise the larger portion of the total error in most of the surface forecast variables that were evaluated. The error characteristics for both surface and upper-air forecasts vary widely by parameter, season and station location. At upper levels, a few characteristic biases are identified. Overall however, the upper-level errors are more nonsystematic in nature and could be explained partly by observational measurement uncertainty. With a few exceptions, the upper-air results also indicate that 24-h model error growth is not statistically significant. In February and August 1997, NCEP implemented upgrades to the eta model's physical parameterizations that were designed to change some of the model's error characteristics near the surface. The results shown in this paper indicate that these upgrades led to identifiable and statistically significant changes in forecast accuracy for selected surface parameters. While some of the changes were expected, others were not consistent with the intent of the model updates and further emphasize the need for ongoing sensitivity studies and localized statistical verification efforts.

Objective verification of point forecasts is a stringent measure of model performance, but when used alone, is not enough to quantify the overall value that model guidance may add to the forecast process. Therefore, results from a subjective verification of the meso-eta model over the Florida peninsula are discussed in the companion paper by Manobianco and Nutter. Overall verification results presented here and in part II should establish a reasonable benchmark from which model users and developers may pursue ongoing eta model verification strategies in the future.

1. Introduction

For several years, Model Output Statistics (MOS; Glahn and Lowry 1972, Carter et al. 1989) from models such as the National Centers for Environmental Prediction's (NCEP) Medium Range Forecast and Nested Grid Models have been used prevalently as sources of localized point forecast guidance. Given an adequately populated sample of runs in which the model configuration is not changed, MOS provides added value to the forecast process by statistically accounting for characteristic strengths and weaknesses in model forecasts at specific locations. However, NCEP is now entering an era where improvements in modeling capabilities are occurring so rapidly (McPherson 1994) that traditional applications of MOS may no longer be appropriate for newer models. On the other hand, the combination of data assimilation techniques, refinements in model physics, and advances in computing efficiency (McPherson 1994) are enabling the possibility for ever more accurate deterministic model point forecasts.

In order to maximize the benefits of point forecast guidance from newer models within an environment of ongoing changes, it is helpful for both model users and developers to maintain an objective awareness of the model's error characteristics at given locations. For example, the local development of techniques that help correct identifiable model errors in real time could improve objective point forecast accuracy (e.g. Homleid 1995; Stensrud and Skindlov 1996; Baldwin and Hrebenach 1998). Moreover, periodic examination of model error characteristics could help developers diagnose and correct possible deficiencies in the model's physical parameterizations.

In the spring of 1996, the Applied Meteorology Unit (AMU) began an evaluation of the NCEP 29-km eta (meso-eta) model in order to document its error characteristics for the U.S. Air Force 45th Weather Squadron (45WS), the National Weather Service (NWS) Melbourne (MLB), and the NWS Spaceflight Meteorology Group (SMG). The mission of the AMU is to evaluate and transition new technology, tools, and techniques into the real-time operational weather support environment for the NWS MLB, SMG, and 45WS (Ernst and Merceret 1995). The NWS MLB is responsible for publicizing daily regional forecasts and warnings of hazardous weather across east-central Florida (Friday 1994). The 45WS provides forecast and weather

warning support for ground processing and launch operations of the space shuttle and other expendable vehicles primarily at the Kennedy Space Center (KSC), Cape Canaveral Air Station (CCAS), and Patrick Air Force Base in east-central Florida (Boyd et al. 1995; Hazen et al. 1995). Among other responsibilities, SMG provides weather support for normal space shuttle end-of-mission and possible launch abort landing scenarios at locations around the world including KSC and Edwards Air Force Base (EDW), California (Brody et al. 1997).

The objective portion of the meso-eta evaluation was designed to assess the model's point forecast accuracy at three selected locations that are important for NWS MLB, 45WS, and SMG operational concerns. Objective verification of point forecasts is a stringent measure of model performance, but when used alone, is not enough to quantify the overall value that model guidance may add to the forecast process. This is especially true for models with enhanced spatial and temporal resolution that may be capable of generating meteorologically consistent, though not necessarily accurate mesoscale weather phenomena (e.g. Cortinas and Stensrud 1995). With this in mind, the AMU also performed a subjective verification of meso-eta model forecasts to help quantify the added value which cannot be inferred solely from an objective evaluation. Results from the AMU's subjective verification of the meso-eta model over the Florida peninsula are discussed in a companion paper (Manobianco and Nutter 1998).

In this paper, results from the objective component of the meso-eta model verification at Edwards Air Force Base, California (EDW), the Shuttle Landing Facility, Florida (TTS) and Tampa International Airport, Florida (TPA) are discussed. Emphasis is placed on establishing the meso-eta model's basic warm and cool season error characteristics at these three locations and on determining if model updates between the evaluation periods led to statistically significant changes in forecast accuracy. The TTS and EDW stations are selected because they are the primary and secondary landing sites for the Shuttle. The TPA site is chosen to compare model errors at two coastal stations on the eastern (TTS) and western (TPA) edge of the Florida peninsula. Note that model sensitivity tests necessary to isolate the exact sources of forecast errors following Manning and Davis (1997) are beyond the scope of this study. The paper is

organized as follows. A brief overview of the eta model and its configuration is presented in section 2. Procedures for data collection and statistical analysis are described in section 3. Results for surface and upper-air forecasts are presented in sections 4 and 5, respectively, followed by a concluding discussion in section 6.

2. Eta model overview

The primary mesoscale modeling efforts at NCEP are focused on the development of the eta model (Rogers et al. 1995). The original version of the eta model with a horizontal resolution of 80 km and 38 vertical layers replaced the Limited-Area Fine Mesh model in June 1993 (Black 1994). In October 1995, NCEP increased the horizontal resolution of the operational "early" eta model from 80 km to 48 km. At the same time, a cloud prediction scheme (Zhao et al. 1997) was implemented and initial analyses were produced using the Eta Data Assimilation System (Rogers et al. 1996). In August 1995, NCEP also began running a mesoscale version of the eta (meso-eta) model with a horizontal resolution of 29 km and 50 vertical layers (Mesinger 1996). Following model upgrades on 31 January 1996 (Chen et al. 1996; Janjic 1996a; Janjic 1996b; Janjic 1996c; Betts et al. 1997), the "early" and meso-eta model configurations became identical except for resolution and data assimilation procedures. The relevant numerics and physics of the eta model are summarized in Table 1.

NCEP implemented two major changes to the eta model's physical parameterizations during the AMU's objective evaluation period. On 18 February 1997, components of the soil, cloud, and radiation packages were updated in both models [Betts et al. 1997 (hereafter BE97); Black et al. 1997 (hereafter BL97); EMC 1997]. These modifications were designed to help control excessive net shortwave radiation at the ground that led indirectly to a bias in the diurnal range of surface temperatures, excessive mixing of the planetary boundary layer (PBL), and a negative bias in surface dew point temperatures. On 19 August 1997, calculation of the model's PBL depth was adjusted to correct for an underestimation of vertical moisture transport out of the lowest model layers (EMC 1997). A portion of the results shown below indicate that

combined effects of these changes led to identifiable and statistically significant changes in forecast accuracy for a few selected parameters.

3. Data and analysis method

The AMU's objective and subjective verification was originally designed to consider 29-km eta model forecast errors over separate four-month periods from May through August 1996 (warm season) and from October 1996 through January 1997 (cool season). Given the ongoing changes to the eta model configuration and the small sample sizes obtained from these limited four-month verification periods, the objective portion of the evaluation was extended to include secondary warm and cool season periods from May through August 1997 and October 1997 through January 1998, respectively. The correspondence between these twin-seasonal evaluation periods and relevant eta model updates is described in Table 2. The most substantial modifications were implemented in February 1997 at a time that falls between the 1996 and 1997 datasets. The timing of this update is convenient for the identification of changes in forecast accuracy, particularly for variables influenced by boundary layer processes.

Forecasts from the 0300 UTC and 1500 UTC meso-eta model cycles were obtained via the internet from the National Oceanic and Atmospheric Administration's (NOAA) Information Center (NIC) ftp server[1]. These files contain 33-h forecasts of surface and upper air parameters at 1-h intervals. NCEP extracts these surface and upper air station forecasts from the meso-eta model grid point nearest to the existing rawinsonde observation sites.

Hourly surface observations from TTS, TPA, and EDW are used to verify meso-eta point forecasts of 10-m wind speed and 2-m temperature and dew point temperature. Upper-air forecasts of wind speed, temperature, and mixing ratio are verified using available rawinsonde observations from EDW, CCAS (XMR), and Tampa Bay (TBW). Log-linear interpolation of

[1] At the time of writing, meso-eta forecast point and grid data can be obtained via anonymous ftp from nic.fb4.noaa.gov.

data is used between reported pressure levels for verification at 25-mb intervals from 1000 to 100 mb. While surface forecasts are verified hourly, upper-air forecasts are verified only for those hours coinciding with the available rawinsonde release times. Surface and rawinsonde observation sites are not collocated at XMR and TBW, but the available sites are separated by not more than about 30 km (i.e. the meso-eta model grid spacing). In order to avoid confusion, all subsequent references to rawinsonde and surface verification will use the rawinsonde station identifiers XMR, TBW, and EDW.

The statistical measures used to quantify model forecast errors are the bias (forecast – observed), root mean square (RMS) error, and error standard deviation. For interpretation of results, it is helpful to recognize that the total model error includes contributions from both systematic and nonsystematic sources. Systematic errors (model biases) are usually caused by a consistent misrepresentation of such factors as orography, radiation and convection. Nonsystematic errors are indicated by the error standard deviations and represent the random error component caused by initial condition uncertainty or inconsistent resolution of scales between the forecasts and observations. While it could be possible to partially correct for known systematic errors by subtracting the bias, the nonsystematic errors are rather unpredictable in nature and may contribute to a degraded daily forecast product. In order to determine if model updates led to a statistically significant annual change in forecast accuracy, a Z statistic (Walpole and Meyers 1989) is calculated for a given parameter and compared with the normal distribution using a 99% confidence level. Additional details regarding statistical calculations are provided in the Appendix.

For quality control, gross errors in the data are screened manually and corrected, if possible. Errors which are greater than three standard deviations from the mean error (bias) are excluded from the final statistics. This procedure is effective at flagging bad data points and removes less than one percent of the data.

4. Surface results

In the following section, meso-eta point forecast error characteristics for 2-m temperature and dew point temperature and 10-m wind speed are established for both the 1996 warm and cool seasons. Then an examination of errors for the 1997 warm and cool season periods highlights changes in forecast accuracy which may have occurred following the February and August 1997 model upgrades (Table 2). Although statistics were calculated separately for the 0300 and 1500 UTC forecast cycles, only those from the 0300 UTC cycle are shown here. Results from the 1500 UTC cycle provide little additional information since positive or negative biases occur with comparable magnitudes at approximately the same time of day in both forecast cycles. Moreover, averaging data from both the 0300 and 1500 UTC cycles as a function of forecast duration tends to cancel out the diurnally varying errors.

a. 1996 Warm Season

During the 1996 warm season, biases in 2-m temperature at XMR and TBW follow a diurnal cycle as the mean errors range from about –3 to 1 °C (Fig. 1a). The amplitude of the diurnal cycle is larger at EDW, with cold biases reaching almost –6 °C during the early part of the forecast. Since forecast biases and corresponding RMS errors are comparable in magnitude at EDW (Figs. 1a, b), the larger contribution to the total error for this location evidently is derived from a systematic model error. One possible explanation for this apparent model deficiency at EDW may be that the forecast point data extracted from the model are almost 250 m lower than the actual station elevation. The results at all three locations are also consistent with those from BE97 and BL97 who found an excessive range of summer temperatures due to radiation errors in the 1996 version of the 48-km eta model.

Warm season biases in 2-m dew point temperature at XMR and TBW are generally within ±2 °C (Fig. 1d). Biases at EDW are positive during the first 21 h of the forecast cycle (Fig. 1d). When viewed in conjunction with the 2-m temperature bias in Fig. 1a, the net result is that forecasts are too cold and moist over this period. The studies by BE97 (their Fig. 10b) and BL97 (their Fig. 4b) indicate excessive amounts of 2-m specific humidity in the forecasts at time zero

using regionally averaged data during the summer. Their results also reveal that specific humidity levels are underforecast on average throughout the remainder of the forecast cycle. Here, zero-hour dew point errors at EDW are consistent with results from those studies but the enduring positive bias indicates clearly that regionally averaged statistics can mask important error characteristics that are specific to particular locations. Some of the difficulties in forecasting dew point temperatures at EDW could relate to problems with PBL mixing and/or incorrect specification of soil moisture processes as discussed by BE97. Such difficulties would likely be exacerbated by the station elevation error at EDW and also by post-processor errors while translating mixing ratios into 2-m dew point temperatures.

Warm season biases in 10-m wind speed range from 0 to –5 m s^{-1} at EDW and from –1 to 2 m s^{-1} at XMR and TBW (Fig. 1g). Therefore, 10-m wind speed forecasts at XMR and TBW tend to be slightly fast on average while those at EDW are generally too slow. The relatively large increase in the magnitudes of biases and RMS errors at EDW between about 1500 and 0300 UTC reflects a period during which systematic model errors comprise the larger portion of the total forecast error (compare Figs. 1g-i).

b) 1996 cool season

During the 1996 cool season, 2-m temperature biases are slightly positive at XMR and slightly negative at TBW, with errors ranging from about 0 to 2 °C and 0 to -2 °C, respectively (Fig. 2a). Forecast temperatures at EDW are about 0 to -4 °C colder than observed on average. Over the first 12 h of the forecast cycle, large error standard deviations at EDW (Fig. 2c) suggest that nonsystematic errors contribute to a substantial portion of the total model error. During the middle part of the forecast cycle from about 1500 to 0300 UTC, the larger negative bias at EDW indicates that systematic model errors contribute more strongly to the total error. Cool season temperature errors at TBW have nearly the same characteristics as those from the previous warm season. At XMR and EDW however, the cool season results do not clearly show diurnal fluctuations that would otherwise be consistent with an excessive range of temperatures in the

1996 eta model configuration. Additional sensitivity studies are therefore necessary in order to determine other possible sources of systematic model error during the cool season.

Cool season biases in 2-m dew point temperature at all three stations are generally larger than those of the previous warm season (compare Figs. 1d, 2d). Biases at TBW range from about –1 to 3 °C while at XMR, a moist bias of 3 to 4 °C is evident throughout much of the forecast cycle. Qualitatively, the difference in error characteristics at XMR and TBW is notable given their relative proximity. Model biases at EDW follow similar fluctuations with time during both seasons, but reach slightly higher maximum values of around 6 °C during the cool season at 2100 UTC. Difficulties remain at EDW during the cool season for initializing the zero-hour dew point temperatures. The overall cool season increase in forecast biases contributes to a corresponding growth in RMS error at all three locations (Fig. 2e). This result suggests that systematic errors in eta model dew point temperature forecasts are larger during the cool season.

Cool season wind speed biases at XMR are about 1 m s^{-1} greater than those during the warm season (compare Figs. 2g, 1g). The combined cool season increases in forecast biases and error standard deviations at XMR result in RMS errors which are about 1.5 m s^{-1} larger than corresponding warm season errors (Figs. 2h, i). Wind speed biases at TBW are comparable during both seasons while the slow bias at EDW improves in the cool season.

c) 1997 warm season

Model biases during 1997 are presented in order to provide an updated assessment of forecast accuracy following the February 1997 model updates (Table 2). Annual changes in the magnitude of model biases $\left(\left| \overline{\Phi'_{97}} \right| - \left| \overline{\Phi'_{96}} \right| ; \text{see Appendix} \right)$ are also examined to determine whether the model updates improved or degraded the model's systematic errors at XMR, TBW, and EDW. By subtracting the absolute value of the biases, it becomes clear whether model errors increased or decreased on average during 1997. For those instances where the Z scores (see Appendix) reveal statistically significant changes in bias, efforts are made to determine whether the changes may be attributed to annual differences in either the mean forecasts or observations.

Wherever bias changes are explained largely by differences in mean forecast values, it is likely that the model updates led to an improvement or degradation in forecast accuracy during 1997.

During the 1997 warm season, 2-m temperature biases at XMR and TBW range from about 1 to –3 °C while at EDW, forecasts are on average 2 to 6 °C colder than observed throughout much of the forecast cycle (Fig. 3a). Forecast biases at XMR and TBW remain within about 1 °C of their 1996 values (Fig. 3b). At EDW, 2-m temperature biases increase in magnitude by about 3 °C between 1500 and 0600 UTC. The standardized *Z* statistic indicates that these larger errors at EDW are statistically significant at the 99% confidence level (Fig 3c). Average forecast temperatures at XMR and TBW increase slightly during 1997 while those at EDW are reduced by about –3 °C (Fig. 4a). Observed temperature climatologies are nearly identical at all three locations during both 1996 and 1997 (Fig. 4b). These results confirm that the stronger cold bias at EDW during the 1997 warm season is driven mostly by a reduction in forecast temperatures and is therefore consistent with the intent of the February 1997 model updates. Curiously, the further reduction of forecast temperatures at EDW during 1997 contributes to a loss of accuracy for that location. One possible explanation may be that the eta model radiation errors noted by BE97 and BL97 inadvertently masked a more serious error related to an inaccurate specification of the true surface elevation.

The 2-m dew point temperature biases at XMR and TBW maintain a steady value around -1 °C during the 1997 warm season (Fig. 3d). Results at EDW continue to indicate a large positive (moist) bias in the forecasts at time zero and from about 1500 to 0300 UTC. During 1997, the magnitude of the biases at XMR and TBW remain within ±1 °C of their corresponding 1996 values (Fig. 3e). At EDW however, absolute mean errors are diminished over the first part of the forecast cycle, but then are followed by a period where errors increase by nearly 6 °C. The Z statistics shown in Fig. 3f confirm that the annual changes in 2-m dew point temperatures are statistically significant during the middle of the forecast cycle at all three stations. The results shown in Fig. 4c indicate that these annual changes in bias are driven mostly by an increase

(decrease) in the mean forecast values at EDW (XMR and TBW). By comparison, relatively minor shifts are noted in the average dew point temperature observations (Fig. 4d).

The eta model updates implemented in February 1997 were designed to reduce PBL mixing and thereby improve the summer dry bias noted in specific humidity forecasts (BE97; BL97). Although increased values for 2-m dew point temperature forecasts at EDW (Fig. 4c) are consistent with the intent of these model updates, the associated raise in the existing positive (moist) bias during 1996 (Fig. 1d) leads to a further loss of forecast accuracy at that location in 1997 (Fig. 3d, e). Conversely, the lower dew point temperature forecasts at XMR and TBW (Fig. 4c) are not anticipated and lead to annual changes in bias which, although statistically significant, do not clearly represent an improvement in accuracy (Fig. 3d).

Biases for 1997 warm season forecasts of 10-m wind speed at all three stations are nearly identical to those of the 1996 warm season. Again, forecasts tend to be slightly fast at XMR and TBW with a continued slow bias at EDW (Fig. 3g). The only statistically significant annual changes occur at EDW around 0200 UTC where the magnitude of the biases increase by 2 m s^{-1} during 1997 (Figs. 3h-i). It is unclear whether these changes in bias are driven by model updates alone since the differences between 1996 and 1997 mean forecasts and observations are both small (Figs. 4e, f). This result is not surprising since the eta model updates implemented in February 1997 were not designed explicitly to alter the forecast wind fields.

d) 1997 cool season

During the 1997 cool season, 2-m temperature forecasts at XMR are on average about 1 °C warmer than observed (Fig. 5a). At EDW, forecasts are again colder than observed throughout much of the forecast cycle, especially from about 1500 to 0300 UTC. Biases at TBW indicate that the diurnal range of 2-m temperatures is overforecast slightly with values ranging from about −1 to 2 °C. The overall forecast accuracy at XMR and EDW is comparable during both 1996 and 1997 cool seasons as temperature biases remain within about ±1 °C (Fig. 5b). At TBW however, the magnitude of the biases increase by about 2 °C during the middle of the forecast period in 1997. The statistical significance of these annual changes in bias at TBW is supported by the *Z*

statistics shown in Fig. 5c. The increases in mean forecast temperatures at TBW during 1997 are larger than those which occur in the observations (Figs. 6a, b). But an increase in mean forecast temperatures during local daytime hours is not consistent with the intent of the February or August 1997 model updates and actually leads to a degradation of forecast accuracy at TBW. Notably, temperature biases at EDW are nearly identical during both cool seasons whereas annual differences in warm season data suggest a strong response to the February 1997 model updates.

Biases in 2-m dew point forecasts at XMR and TBW are quite good on average during the 1997 cool season (Fig. 5d). However, results at EDW continue to indicate a large positive (moist) bias in the forecasts at time zero and also during the latter portions of the forecast cycle. While dew point temperature biases at EDW are similar during both 1996 and 1997 cool seasons, the magnitudes of the errors at XMR and TBW decrease by about 3 °C (Fig. 5e). The *Z* statistic reveals that the improvements noted at XMR and TBW are statistically significant at the 99% confidence level (Fig. 5f). The enhanced forecast accuracy at XMR and TBW evidently results from a combination of lower (drier) dew point temperature forecasts and higher (wetter) observations on average during 1997 (Figs. 6c, d). In spite of the forecast improvements, these findings are not anticipated since the February and August 1997 model updates were designed to raise near-surface moisture levels (BE97; BL97). The statistics shown here provide further evidence that additional sensitivity studies are needed to help identify other possible sources of systematic model error during the cool season.

Wind speed forecasts at all three stations during the 1997 cool season are nearly identical to their 1996 values (compare Figs. 2g, 5g). The greatest annual changes in the magnitude of the biases occur at XMR where forecast accuracy improves by at most 1 m s^{-1} (Fig. 5h). The Z scores shown in Fig. 5i confirm that no statistically significant changes occur in wind speed biases between the 1996 and 1997 cool seasons. Moreover, differences in mean forecast and observed wind speed between 1996 and 1997 are similar (Figs. 6e, f). As during the warm season, this cool season result is expected since the eta model updates were not explicitly designed to modify wind speed forecasts.

5. Upper air results

Examination of warm and cool season error statistics for upper-air forecasts of temperature, mixing ratio and wind speed at XMR, TBW, and EDW reveals only subtle changes in their characteristics between 1996 and 1997 (annually stratified results not shown). This is not surprising since the eta model changes implemented in February and August 1997 (Table 2) were designed primarily to improve documented deficiencies in forecasts for surface and boundary layer variables (BE97, EMC 1997). For these reasons, no attempt is made here to identify statistically significant annual changes in the upper-air forecast errors and attribute them solely to eta model updates. Instead, all data collected during 1996 and 1997 are pooled into their respective warm and cool season periods to develop generalized profiles of meso-eta error characteristics at XMR, TBW, and EDW.

a) Temperature

Warm season temperature biases at EDW are less than ±1 °C (Fig. 7a). At XMR and TBW, forecast biases below 700 mb are about 1 °C colder than observed whereas above 700 mb they are about 1 to 2 °C warmer than observed. The net effect for warm season forecasts at the Florida stations is a tendency towards a thermally stable model atmosphere. RMS errors range from about 1 to 2.5 °C and are largest in the upper troposphere (Fig. 7b). In comparison, typical RMS uncertainty in rawinsonde temperature observations is about 0.6 °C (Hoehne 1980; Ahnert 1991). This fact suggests that about half the nonsystematic error between the forecasts and observations may be due to measurement uncertainty.

During the cool season, temperature forecasts at EDW exhibit a negative (cold) bias below 700 mb that exceeds -4 °C near the surface (Fig. 7d). At XMR and TBW, temperature errors are less than 1 °C except around the 700 mb level and above the tropopause. Examination of individual forecast and observed soundings at XMR throughout the cool season (not shown) reveals that the 700 mb cold bias appears primarily because model forecasts of the lower tropospheric inversion height are frequently at a higher level than where they are actually observed. In the middle troposphere, RMS errors in cool season temperature forecasts at EDW

are substantially larger those at XMR and TBW (Fig. 7e). Since biases are small above 700 mb at EDW, the relatively large error standard deviations suggest that a greater portion of the total RMS error is caused by a large amount of day-to-day variability in the forecast errors (Fig. 7f).

b) Mixing ratio

Warm season mixing ratio biases at XMR and TBW (Fig. 8a) indicate that meso-eta forecasts are on average about 1 g kg^{-1} too dry below 700 mb. Conversely, mixing ratio biases at EDW are about 0.5 g kg^{-1} greater than observed. Between 700 and 500 mb, forecasts at all three locations indicate a positive (dry) bias while above 500 mb they tend to retain excessive amounts of moisture. In combination with the negative lower tropospheric temperature biases discussed previously, these results suggest that warm season model forecasts at XMR and TBW are typically more stable than observed. Cool season mixing ratio biases at all three locations reveal excessive moisture near the surface with a rapid vertical transition to a layer with less moisture than observed (Fig. 8d).

RMS errors for the warm season (Fig. 8b) drop from around 2.5 g kg^{-1} at low-levels (1.5 g kg^{-1} at EDW) to near zero at 200 mb, where there is very little water vapor present in the atmosphere. In the cool season, RMS errors follow a similar profile at all three stations starting with values of 2 g kg^{-1} near the surface (Fig. 8e). Since the error standard deviations shown in Figs. 8c and 8f are more than double the magnitude of the mixing ratio biases, nonsystematic errors account for roughly 50 to 75% of the total RMS error. Results shown in Figs. 8b and 8e are consistent with those of Rogers et. al (1996), who show 24-h RMS errors in specific humidity from 48-km eta model forecasts across the United States during September 1994 ranging from nearly 2 g kg^{-1} at 1000 mb to less than 0.1 g kg^{-1} at 250 mb (see their Fig. 7). Note that these calculations for mixing ratio errors are not normalized by magnitude and are therefore not representative of percent errors as the mixing ratio tends toward zero in the upper troposphere.

c) Wind speed

Warm season wind speed biases are generally less than ±1 m s^{-1} (Fig. 9a). The exception occurs at EDW where lower tropospheric wind speed forecasts are about 2 m s^{-1} slower than observed. This result is consistent with the negative (slow) bias in 10-m wind speed forecasts identified at EDW (Fig. 1g). Below 400 mb, warm season RMS errors range from about 2 to 4 m s^{-1} (Fig. 9b). RMS errors around the 200 mb level are larger with values approaching 6 m s^{-1}. Since forecast biases are small and uncertainties in rawinsonde wind speed measurements are about 3.1 m s^{-1} (Hoehne 1980), much of the total RMS wind speed error at lower levels could result from observational uncertainty.

During the cool season, forecast wind speeds at XMR and TBW are about 1 m s^{-1} slower (faster) than observed in the middle (upper) troposphere (Fig. 9d). At EDW, wind speed biases range from 1 to 3 m s^{-1} except near the surface where forecast wind speeds remain slow. Cool season RMS errors at XMR and TBW are comparable to those found during the warm season and again, likely receive large contributions from observational measurement uncertainties (Fig. 9e). At EDW, cool season RMS errors above 700 mb are nearly double those of the warm season with increased contributions from both systematic and nonsystematic errors.

d) Forecast error growth

Since rawinsonde observations are available only twice daily under normal circumstances, it is not possible to observe the temporal evolution of upper level forecast errors on an hourly basis throughout the forecast cycle[2] . However, separate examination of seasonal forecast errors at three 12-h intervals (not shown) reveals that upper-level errors do fluctuate slightly with forecast duration although their vertical profiles remain qualitatively similar. Unlike the surface error characteristics, diurnal oscillations are not evident in the upper-air forecast errors above the lowest few levels. A paired Z statistic is used therefore to determine if seasonal mean changes in

[2] The 50 MHz wind profiler data at KSC/CCAS are available every 5 min but are not used for the objective portion of this study because similar data are not available at TBW or EDW.

upper-level model biases during a 24-h period represent a statistically significant systematic error growth (see Appendix).

Examination of statistics at each of the three 12-h verification intervals (not shown) reveals that the lower tropospheric cold bias in forecast temperatures (e.g. Fig 7) at XMR and TBW becomes more negative with time during both warm and cool seasons. The corresponding paired *Z* statistic (Figs. 10a, d) indicates that this growing cold bias is statistically significant over a 24-h period. Following a similar argument, the positive bias in upper tropospheric temperature forecasts during the warm season at TBW tends to grow stronger with time (Fig. 10a). The significance of the warm season error growth in forecast temperatures near the surface at EDW (Fig. 10a) is questionable due to the possible influence of diurnal variability (e.g., Fig. 1a). Most error growth in mixing ratio forecasts is not statistically significant except near the 200 mb level at TBW (Figs. 10b, e). The only significant change in systematic error (bias) for wind speed forecasts is found near the tropopause during the warm season at XMR and TBW (Fig. 10c). Again, examination of statistics at each of the three 12-h verification intervals (not shown) reveals that the positive (fast) bias in upper tropospheric wind speed forecasts at these locations (e.g. Fig 9a) tends to diminish with time during the warm season. Although a few exceptions are noted here, the results shown in Fig. 10 reveal that the mean 24-h error growth for temperature, mixing ratio, and wind speed is not statistically significant at the 99% confidence level.

6. Summary and discussion

From May 1996 through January 1998, the AMU conducted warm- and cool-season evaluations of meso-eta surface and upper-air point forecast accuracy at XMR, TBW, and EDW. These three locations were selected because they are important for 45WS, NWS MLB, and SMG operational concerns. Each warm- and cool-season verification period extends from May through August and October through January, respectively. By extending the evaluation for a second consecutive year, it was possible to identify statistically significant changes in the error characteristics which developed in response to the February and August 1997 model updates

(BL97; EMC 1997). The twin-season comparison of forecast accuracy is helpful for model users by highlighting the model's characteristic strengths and weaknesses before and after the incorporation of model updates. Such results are also helpful for model development efforts and emphasize the need for ongoing analysis of model errors at specific locations.

During the 1996 warm season, forecast errors in 2-m temperature followed a diurnal cycle that was most pronounced at EDW. This excessive range of temperatures is consistent with results documented by BE97 and BL97 using regionally averaged data during the summer. In contrast, the strong positive (moist) bias in 2-m dew point temperature forecasts at EDW is not in agreement with the negative (dry) bias identified by BE97 and BL97. Finally, the 10-m wind speed forecasts at XMR and TBW tended to be slightly fast on average while those at EDW were generally too slow.

The 2-m temperature forecast biases during the 1996 cool season have similar magnitudes as those during the warm season, but do not clearly show the diurnal fluctuations that would otherwise be consistent with an excessive range of forecast temperatures. Cool season biases in 2-m dew point temperature forecasts are mostly positive (moist) at all three locations and larger in magnitude than compared to the previous warm season. The positive (fast) bias in 10-m wind speed forecasts at XMR increases during the cool season while at EDW the negative (slow) bias is diminished.

The model updates implemented in January and August 1997 (Table 2) were based on regional verification of forecasts from parallel operational and test runs of the 48-km eta model during the warm season (BL97; EMC 1997). The updates were designed primarily to improve noted deficiencies in point forecasts of low-level temperature and moisture. The 1996 results shown here for the 29-km eta model reveal that warm season dew point temperature biases at EDW and most cool season variables at XMR, TBW and EDW are not consistent with the regional verifications that led to the implementation of the model updates. While differences in resolution between the 29- and 48-km eta models could account for some of the discrepancies, the use of regionally averaged data can mask specific localized deficiencies especially when

including data from areas of varying elevation (Manning and Davis 1997). By comparing statistical results from the 1996 warm and cool seasons with those during 1997, it is possible to determine if the model updates produced the intended improvements in forecast accuracy specifically at XMR, TBW, and EDW.

During the 1997 warm season, 2-m temperature forecast biases at XMR and TBW remained within about 1 °C of their 1996 values. At EDW, forecasts were on average about 2 to 6 °C colder than observed throughout the forecast cycle. The stronger cold bias at EDW was identified as a statistically significant increase in error. Although the decrease in average forecast temperature is consistent with the eta model updates implemented in February 1997, the change ultimately led to an overall loss of forecast accuracy for that location. The large positive (moist) bias evident at EDW during the 1996 warm season grew even larger in 1997. Again, this increase in moisture is expected, but resulted in forecasts of lower quality on average. The general increase in forecast errors at EDW between 1996 and 1997 suggests that the eta model radiation errors noted by BE97 and BL97 may have inadvertently masked an additional error related to an inaccurate specification of the true surface elevation at that location. Wind speed biases in 1997 remained nearly identical to their corresponding values from the previous year.

The only changes in 2-m temperature biases during the 1997 cool season that were considered statistically significant occurred during the middle of the forecast cycle at TBW. The increase in average forecast temperature which occurred there resulted in diminished forecast accuracy and was not anticipated given that the February 1997 model update was designed to help reduce daytime temperatures. The 2-m dew point temperature errors at EDW were similar during both the 1996 and 1997 cool seasons, but statistically significant improvements in forecast biases were found at XMR and TBW. Again, the reduction of average forecast dew point temperatures at XMR and TBW during 1997 which led to these improvements was not expected. Wind speed forecasts during the 1997 cool season remained nearly unchanged relative to their corresponding values in 1996.

Interpretation of results from the twin-season verification is rather complicated since forecast biases showed improvement (or degradation) in many cases irrespective of whether such changes were consistent with the intent of the February and August 1997 model updates. In general, systematic model deficiencies comprise the larger portion of the total error in forecasts of most surface variables considered here. The errors for the surface parameters described here were generally smallest at TBW and largest at EDW. Similarly, annual changes in error characteristics remained the most stable at TBW while annual changes at EDW responded strongly to model updates. While some of the results were expected, others were not consistent with the intent of the model updates and further emphasize the need for additional sensitivity studies and ongoing localized verification.

In addition to the surface forecast verification, the AMU conducted an objective verification of upper-level point forecast accuracy at XMR, TBW, and EDW. Results from the upper-air verification did not reveal annual changes in forecast errors that could be attributed solely to the February and August 1997 model updates. In addition, 24-h hour error growth in the upper-level forecasts was not statistically significant except for the few exceptions discussed in section 5. Given these results, the data were pooled together to develop vertical profiles of forecast error characteristics during both warm and cool seasons. Warm season forecasts at XMR and TBW were generally drier and more stable than observed. The height of the lower tropospheric inversion at XMR and TBW was misrepresented during the cool season. The height of the tropopause was also misrepresented during both seasons at all three locations. Errors in wind speed forecasts were reasonably small, but could be explained largely by nonsystematic error such as rawinsonde measurement uncertainty. Although reliable on average, the relatively larger degree of nonsystematic error in the forecasts for most upper-air variables considered here provides evidence of substantial day-to-day variability. Given this variability, real-time assessment of forecast accuracy is necessary on any given day to help users determine if the model forecasts are consistent with current observations.

It is important that forecasters maintain an ongoing awareness of model updates and the effects that such changes will have on point forecast accuracy within their area of responsibility. While model updates are generally well tested and designed to improve forecast accuracy, the results shown here demonstrate that the desired effects do not always yield the expected improvements at every location. This may become particularly true as national-scale, operational models with greater horizontal and vertical resolutions are able to forecast explicitly the complex processes that occur within the PBL. In recent years, information documenting model updates has been made available regularly on the internet. Indeed, much of the information needed for writing this paper and maintaining an understanding the model changes was obtained from an internet FAQ written expressly for this purpose (EMC 1997). As forecasters discover localized model deficiencies through ongoing real-time statistical verification strategies, results should be documented regularly and shared with model developers. As expressed by Manning and Davis (1997), "These statistics would provide additional information to model users and alert model developers to those research areas that need more attention". The additional and complementary need for subjective verification strategies in mesoscale models is discussed in the companion paper (Manobianco and Nutter 1998).

On 9 February 1998, NCEP upgraded the horizontal resolution of the "early" eta model from 48 to 32 km with an increase in vertical resolution from 38 to 45 levels (Rogers et al. 1997). In addition, a three-dimensional variational analysis scheme was implemented along with the use of a "partial" continuous Eta Data Assimilation Cycle. Aside from the differences in data assimilation methods at the time of this writing, this version of the "early" eta is similar in resolution and dynamics to the meso-eta model version evaluated here. Therefore the objective verification results presented here for XMR, TBW and EDW and the subjective verification results presented in part II should establish a reasonable benchmark from which model users and developers may pursue ongoing eta model verification strategies in the future.

Acknowledgments

The authors thank Dr. F. Merceret (NASA KSC), Mr. R. Lafosse (SMG), Mr. T. Oram (SMG), and Mr. William P. Roeder (45WS) for their comments and suggestions on an earlier version of the manuscript.

Appendix

The statistical measures used here to quantify model forecast errors are the bias, root mean square (RMS) error, and standard deviation. If Φ represents any of the parameters under consideration for a given time and vertical level, then forecast error is defined as $\Phi' = \Phi_f - \Phi_o$ where the subscripts *f* and *o* denote forecast and observed quantities, respectively. Given N valid pairs of forecasts and observations, the bias is computed as

$$\overline{\Phi'} = \frac{1}{N}\sum_{i=1}^{N}\Phi_i' \qquad \text{(A.1)},$$

the RMS error is computed as

$$\text{RMSE} = [\text{MSE}]^{1/2} = \left[\frac{1}{N}\sum_{i=1}^{N}(\Phi_i')^2\right]^{1/2} \qquad \text{(A.2)},$$

and the standard deviation of the errors is computed as

$$\sigma' = \left[\frac{1}{N}\sum_{i=1}^{N}(\Phi_i' - \overline{\Phi'})^2\right]^{1/2} \qquad \text{(A.3)}.$$

In equation (A.3), N is used rather than N-1 so that a decomposition following Murphy (1988, Eq. 9) could be applied to the MSE:

$$\text{MSE} = \overline{\Phi'}^2 + \sigma'^2 \qquad \text{(A.4)}.$$

Therefore, the total model error consists of contributions from model biases ($\overline{\Phi'}^2$) and random variations in the forecast and/or observed data (σ'^2). Note that if the model bias or systematic error is small, most of the MSE is due to random, nonsystematic type variability in the errors. Murphy's (1988) decomposition of the MSE considered individually the error contributions from the model bias and from the sample variances and covariance of the forecasts and observations. Here, Eq. A.4 represents an algebraic simplification of that decomposition and quantifies the portion of the MSE that is due to the bias and the variance of the forecast *errors*.

Tests are applied to the surface data in order to determine if model updates led to statistically significant changes in mean forecast error between the 1996 and 1997 warm and cool season periods. Following the Central Limit Theorem as described in most statistical texts, it is

assumed that the sampling distribution for the difference in mean forecast error between 1996 and 1997 is approximately normal. Sample sizes of O(100) for each season enable use of the standardized Z statistic where

$$Z = \frac{\overline{\Phi'_{97}} - \overline{\Phi'_{96}}}{\left\{ \delta_{96}\left[(\sigma'_{96})^2/N_{96} \right] + \delta_{97}\left[(\sigma'_{97})^2/N_{97} \right] \right\}^{1/2}} \qquad \text{(A.5)},$$

the variance inflation factor, $\delta = (1 + \rho)/(1 - \rho)$, and ρ is the lag-1 day autocorrelation for each seasonal time series of data. The variance inflation factor helps prevent the overestimation of Z by adjusting the variance of the sampling distribution to account for the influence of serial dependence, or day-to-day persistence, within the seasonal time series average (Wilks 1995). A two-tailed comparison of Z to the normal distribution using a 99% confidence level has critical values of ±2.58 (Walpole and Meyers 1989). Calculated values of Z that lie outside this critical range indicate that the data are able to support a statistically significant difference between the 1996 and 1997 seasonal mean forecast errors.

The statistical significance of upper-level systematic error growth from early to latter stages of the forecast cycle is determined using a paired Z statistic. The paired Z statistic normalizes the seasonally averaged difference in forecast error between two times during the ith cycle by the associated sample standard deviation. The covariance between errors in the early and latter stages of the forecast is included because the parameters from the ith cycle are not independent and do not necessarily have equal variances (Walpole and Meyers 1989). Here, the paired Z statistic is denoted by Z' where

$$Z' = \frac{\sum_{i=1}^{N} \left(\Phi'_{2i} - \Phi'_{1i} \right)}{\left\{ N\left[(\sigma'_1)^2 + (\sigma'_2)^2 - (\sigma'_{12})^2 \right] \right\}^{1/2}} \qquad \text{(A.6)}.$$

The subscripts 1 and 2 denote variables from the ith forecast cycle verifying at 6-9 h and 30-33 h, respectively. The times used for verification are separated by 24 h and are taken at forecast durations that vary slightly according to balloon release times. Other notations are as above except that $(\sigma'_{12})^2$ denotes the sample covariance. Again using a 99% confidence level, values of

Z' that lie outside the critical values of ±2.58 indicate that the data are able to support a statistically significant 24-h systematic error growth in the upper-air forecasts.

References

Ahnert, P. R., 1991: Precision and compatibility of National Weather Service upper air measurements. Preprints, *Seventh Symp. on Meteorological Observations and Instrumentation,* New Orleans, LA, Amer. Meteor. Soc., 221-226.

Baldwin, M. E., and S. D. Hrebenach, 1998: Experiments with bias-corrected temperature guidance using NCEP's mesoscale eta model. Preprints, *Sixteenth Conf. on Weather Analysis and Forecasting,* Phoenix, AZ, Amer. Meteor. Soc., 388-389.

Betts, A. K., F. Chen, K. E. Mitchell, and Z. I. Janjic, 1997: Assessment of the land surface and boundary layer models in two operational versions of the NCEP eta model using FIFE data. *Mon. Wea. Rev.*, **125**, 2896-2916.

Black, T. L., 1994: The new NMC mesoscale eta model: description and forecast examples. *Wea. Forecasting*, **9**, 265-278.

——, M. Baldwin, K. Brill, F. Chen, G. DiMego, Z. Janjic, G. Manikin, F. Mesinger, K. Mitchell, E. Rogers, and Q. Zhao, 1997: Changes to the Eta forecast systems. *NWS Technical Procedures Bulletin 441*, NOAA/NWS. [Available from National Weather Service, Office of Meteorology, 1325 East-West Highway, Silver Spring, MD 20910.]

Brody, F. C., R. A. Lafosse, D. G. Bellue, and T. D. Oram, 1997: Operations of the National Weather Service Spaceflight Meteorology Group. *Wea. Forecasting*, **12**, 526-544.

Boyd, B. F., W. P. Roeder, J. B. Lorens, D. S. Hazen, and J. W. Weems, 1995: Weather support to pre-launch operations at the Eastern Range and Kennedy Space Center. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 135-140.

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting*, **4**, 401-412.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H.-L. Pan, V. Koren, Q. Duan, and A. Betts, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE results. *J. Geophys. Res.*, **101,** 7251-7268.

Cortinas, J. V., Jr., and D. J. Stensrud, 1995: The importance of understanding mesoscale model parameterization schemes for weather forecasting. *Wea. Forecasting*, **10,** 716-740.

EMC, cited 1997: Environmental Modeling Center: Mesoscale Modeling Branch FAQ. [Available on-line from http://nic.fb4.noaa.gov:8000/research/FAQ-eta.html.]

Ernst, J. A. and F. J. Merceret, 1995: The Applied Meteorology Unit: A tri-agency applications development facility supporting the space shuttle. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 266-269.

Friday, E. W., Jr. 1994: The modernization and associated restructuring of the National Weather Service: An overview. *Bull. Amer. Meteor. Soc.*, **75,** 43-52.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11,** 1203-1211.

Hazen, D. S., W. P. Roeder, B. F. Boyd, J. B. Lorens, and T. L. Wilde, 1995: Weather impact on launch operations at the Eastern Range and Kennedy Space Center. Preprints, *Sixth Conf. on Aviation Weather Systems,* Dallas, TX, Amer. Meteor. Soc., 270-273.

Hoehne, W. E., 1980: *Precision of National Weather Service Upper Air Measurements.* NOAA Tech. Memo. NWS T&ED-16.

Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting*, **10,** 689-707.

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122,** 927-945.

——, 1996a: The Mellor-Yamada level 2.5 turbulence closure scheme in the NCEP Eta model. Research Activities in Atmospheric and Oceanic Modeling, WMO, Geneva, CAS/JSC

WGNE, 4.14-4.15. [Available from World Meteorological Organization, Case Postale 2300, CH-1211 Geneva, Switzerland.]

——, 1996b: The surface layer parameterization in the NCEP Eta Model. Research Activities in Atmospheric and Oceanic Modeling, WMO, Geneva, CAS/JSC WGNE, 4.16-4.17. [Available from World Meteorological Organization, Case Postale 2300, CH-1211 Geneva, Switzerland.]

——, 1996c: The surface layer in the NCEP Eta model. Preprints, *Eleventh Conf. on Numerical Weather Prediction,* Norfolk, VA. Amer. Meteor. Soc., 354-355.

Manning, K. W. and C. A. Davis, 1997: Verification and sensitivity experiments for the WISP95 MM5 forecasts. *Wea. Forecasting,* **12,** 719-735.

Manobianco, J. M., and P. A. Nutter, 1998: Evaluation of the 29-km eta model. Part II: Subjective verification over Florida. Submitted to *Wea. Forecasting.*

McPherson, R. D., 1994: The National Centers for Environmental Prediction: Operational climate, ocean, and weather prediction for the 21st century. *Bull. Amer. Meteor. Soc.,* **75,** 363-373.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.,* **11,** 2637-2649.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.,* **116,** 2417-2424.

Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational "early" eta model: Original 80-km configuration and recent changes. *Wea. Forecasting,* **10,** 810-825.

——, T. L. Black, D. G. Deaven, and G. J. DiMego, 1996: Changes to the operational "early" eta analysis/forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting,* **11,** 391-413.

——, M. Baldwin, T. Black, K. Brill, F. Chen, G. DiMego, J. Gerrity, G. Manikin, F. Mesinger, K. Mitchell, D. Parrish, Q. Zhao, 1997: Changes to the NCEP Operational "Early" Eta Analysis / Forecast System. *NWS Technical Procedures Bulletin 447*, NOAA/NWS. [Available from National Weather Service, Office of Meteorology, 1325 East-West Highway, Silver Spring, MD 20910.]

Stensrud, D. J., and J. A. Skindlov, 1996: Gridpoint predictions of high temperature from a mesoscale model. *Wea. Forecasting*, **11**, 103-110

Walpole, R. E., and R. H. Meyers, 1989: *Probability and Statistics for Engineers and Scientists.* Macmillan, Inc., New York, 765 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, Inc., San Diego, CA, 467 pp.

Zhao, Q., T. L. Black, and M. E. Baldwin, 1997: Implementation of the cloud prediction scheme in the eta model at NCEP. *Wea. Forecasting*, **12**, 697-712.

Figure Captions

Figure 1. Bias, RMS error, and error standard deviation for 2-m temperature and dew point temperature (°C) and 10-m wind speed (m s^{-1}) forecasts from the 0300 UTC meso-eta cycle. Results are plotted for the 1996 warm season as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed).

Figure 2. Same as Fig. 1 but for the 1996 cool season.

Figure 3. 1997 warm season bias, annual difference of absolute bias (AB97 – AB96), and standardized Z statistic for 2-m temperature and dew point temperature and 10-m wind speed forecasts from the 0300 UTC meso-eta cycle. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Units are mks except for the nondimensional Z statistic shown in panels c, f and i. Z scores that lie outside the shaded region indicate that changes between 1997 and 1996 warm season forecast biases are statistically significant at the 99% confidence level (see Appendix).

Figure 4. Comparison of annual changes (1997 – 1996) in mean 0300 UTC meso-eta forecasts and corresponding observations for 2-m temperature and dew point temperature (°C) and 10-m wind speed (m s^{-1}) during the warm season. Results are plotted as a function of time at XMR (solid), TBW (dotted), and EDW (dashed).

Figure 5. Same as Fig. 3 but for the 1997 cool season.

Figure 6. Same as Fig. 4 but for the cool season.

Figure 7. Bias, RMS error, and error standard deviation (°C) of meso-eta temperature forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

Figure 8. Same as Fig. 7 but for mixing ratio (g kg^{-1}).

Figure 9. Same as Fig. 7 but for wind speed (m s^{-1}).

Figure 10. Paired Z statistic plotted as a function of pressure level for meso-eta forecast errors at XMR (solid), TBW (dotted), and EDW (dashed). The nondimensional statistic is shown for temperature (a, d), mixing ratio (b, e) and wind speed (c, f). Warm season values are shown in the left column (panels a-c) while cool season values are shown in the right column (panels d-f). Paired Z values that lie outside the shaded region indicate that 24-h systematic error growth is statistically significant at the 99% confidence level (see Appendix).

Table Captions

TABLE 1. Eta model attributes from Black (1994), Janjic (1994), and Rogers et al. (1996).

TABLE 2. Definition of seasonal verification periods and notable eta model updates.

TABLE 1. Eta model attributes from Black (1994), Janjic (1994), and Rogers et al. (1996).

Dynamics
Model top = 25 mb
Time step = 72 s
Semi-staggered Arakawa E-grid
Gravity wave coupling scheme
Silhouette-mean orography
Split-explicit time differencing
Physics
Explicit grid-scale cloud and precipitation
Modified Betts-Miller convective adjustment
Mellor-Yamada (2.5) for free atmosphere vertical turbulent exchange Mellor-Yamada (2.0) near ground
Geophysical Fluid Dynamics Laboratory radiation scheme
Viscous sublayer over water

TABLE 2. Definition of seasonal verification periods and notable eta model updates.

Verification period	Date began	Date ended	Notable eta model changes (EMC 1997)
1996 warm season	1 May 1996	31 August 1996	
1996 cool season	1 October 1996	31 January 1997	Radiation, cloud fraction, soil moisture, etc. (18 Feb. 1997)
1997 warm season	1 May 1997	31 August 1997	Corrected PBL depth computation (19 Aug. 1997)
1997 cool season	1 October 1997	31 January 1998	

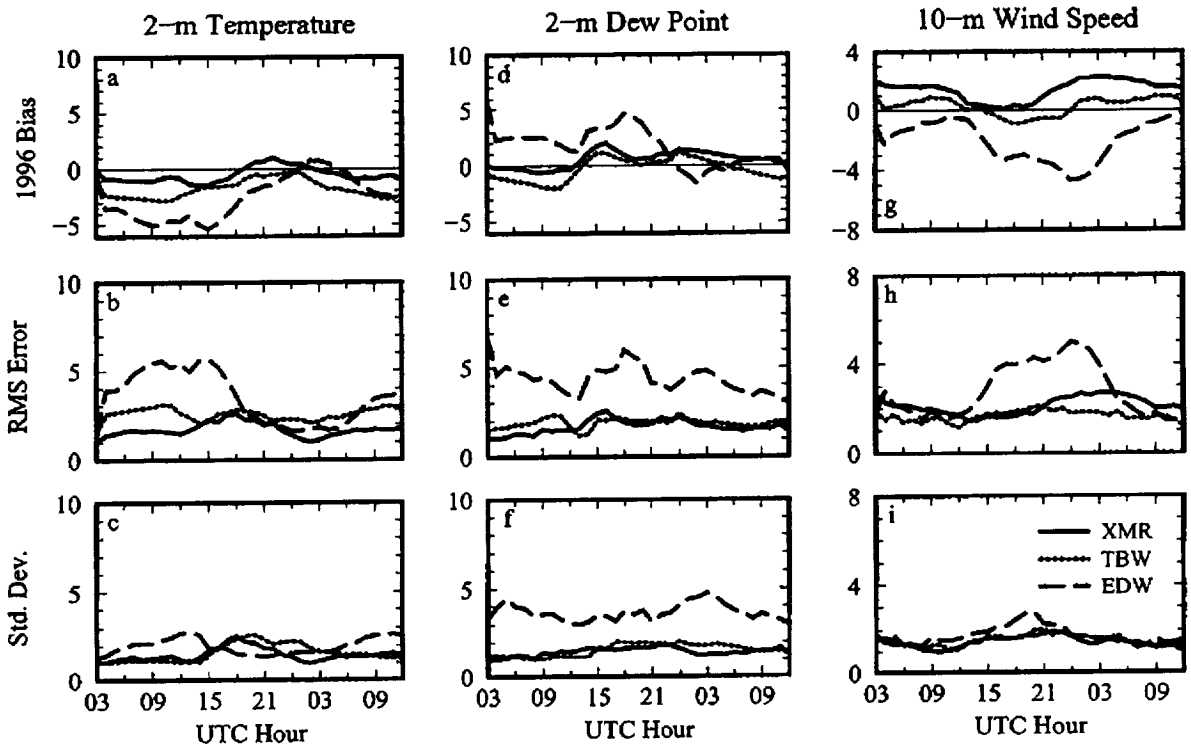

Figure 1. Bias, RMS error, and error standard deviation for 2-m temperature and dew point temperature (°C) and 10-m wind speed (m s-1) forecasts from the 0300 UTC meso-eta cycle. Results are plotted for the 1996 warm season as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed).

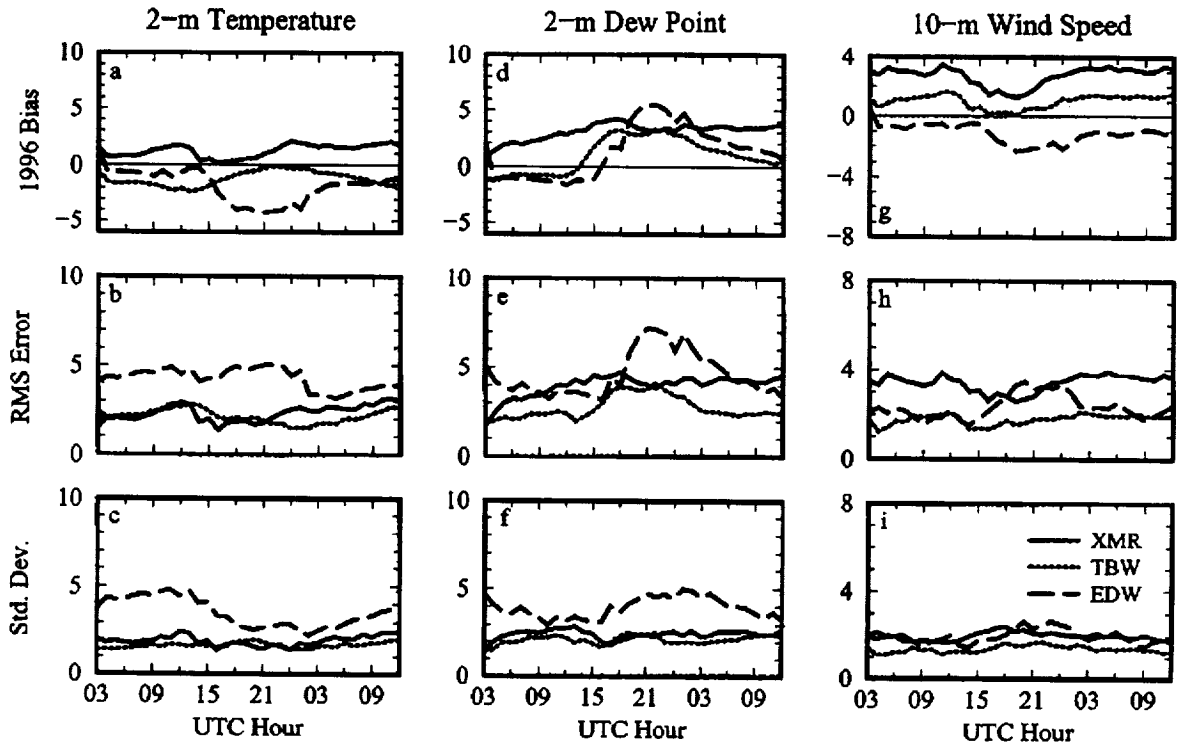

Figure 2. Same as Fig. 1 but for the 1996 cool season.

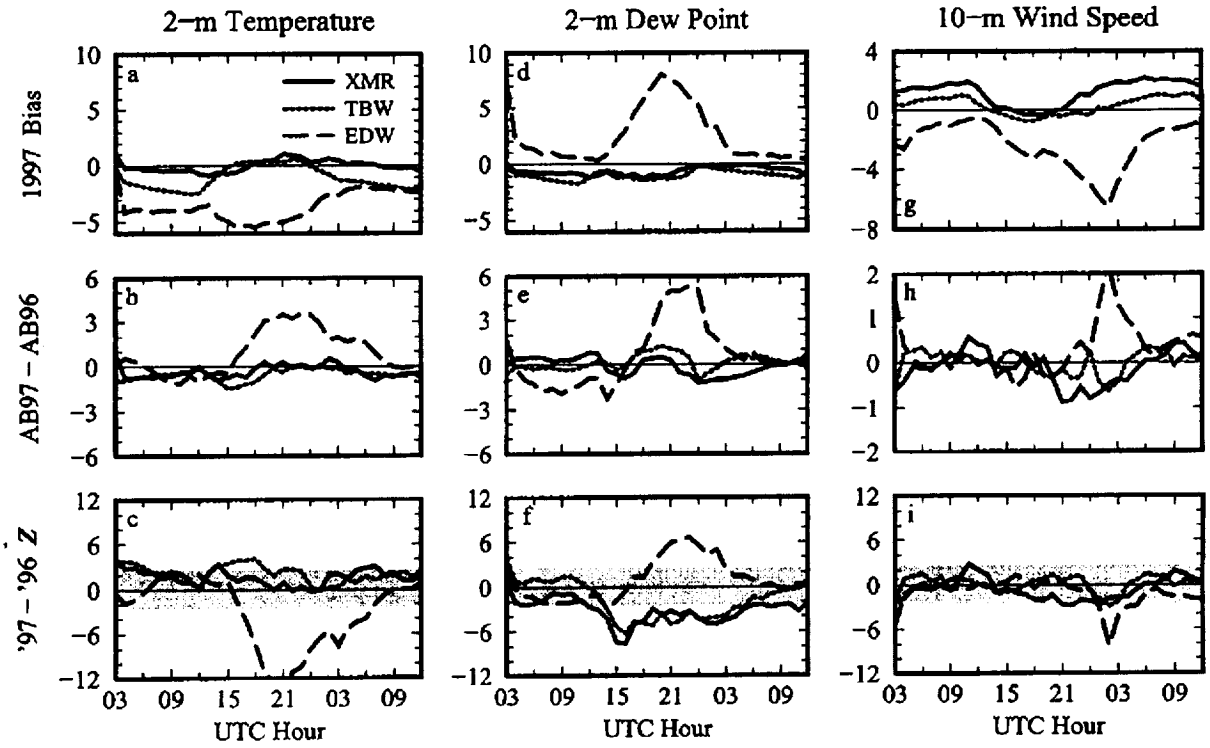

Figure 3. 1997 warm season bias, annual difference of absolute bias (AB97 – AB96), and standardized *Z* statistic for 2-m temperature and dew point temperature and 10-m wind speed forecasts from the 0300 UTC meso-eta cycle. Results are plotted as a function of verification time at XMR (solid), TBW (dotted), and EDW (dashed). Units are mks except for the nondimensional *Z* statistic shown in panels c, f and i. *Z* scores that lie outside the shaded region indicate that changes between 1997 and 1996 warm season forecast biases are statistically significant at the 99% confidence level (see Appendix).

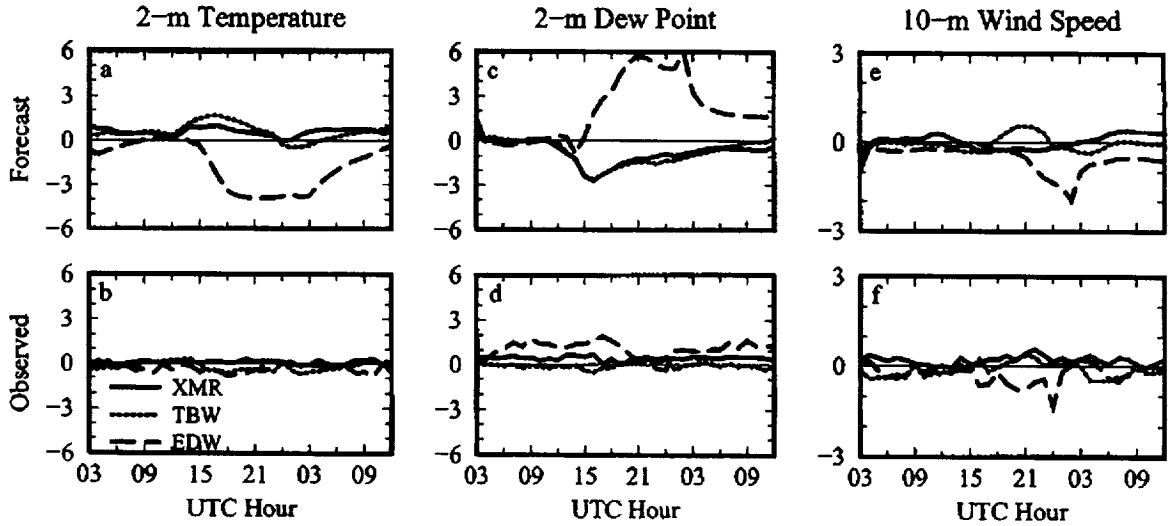

Figure 4. Comparison of annual changes (1997 – 1996) in mean 0300 UTC meso-eta forecasts and corresponding observations for 2-m temperature and dew point temperature (°C) and 10-m wind speed (m s^{-1}) during the warm season. Results are plotted as a function of time at XMR (solid), TBW (dotted), and EDW (dashed).

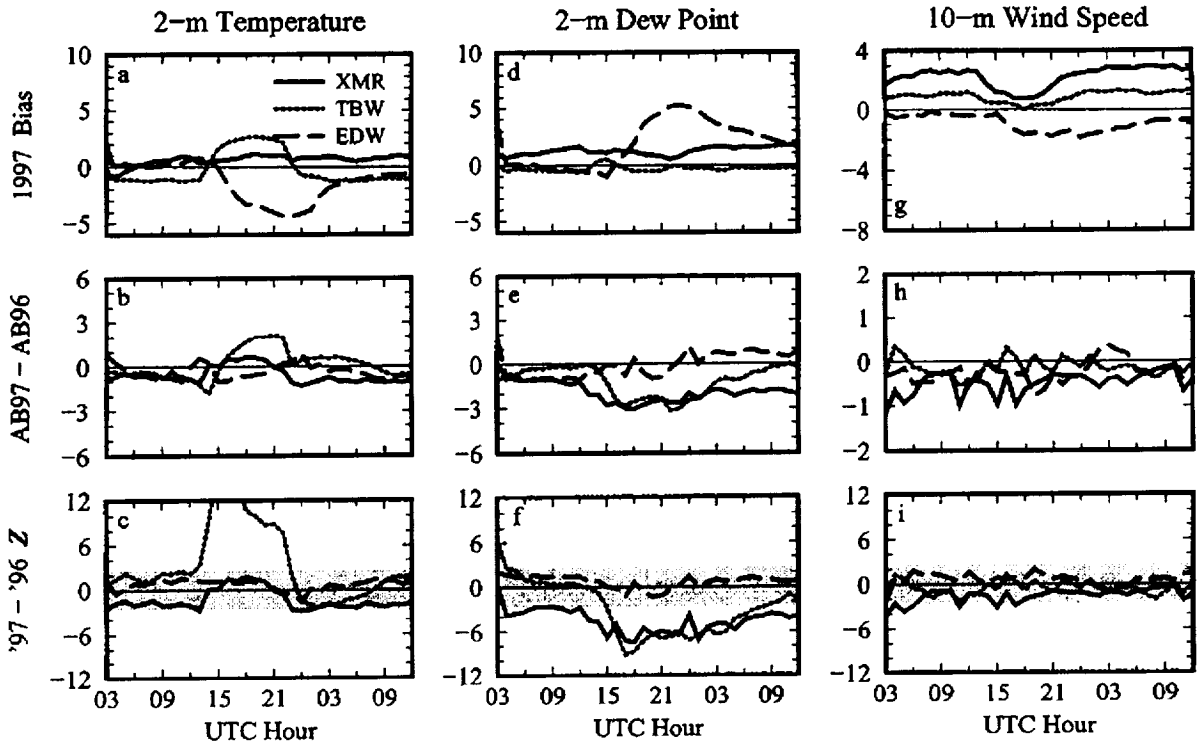

Figure 5. Same as Fig. 3 but for the 1997 cool season.

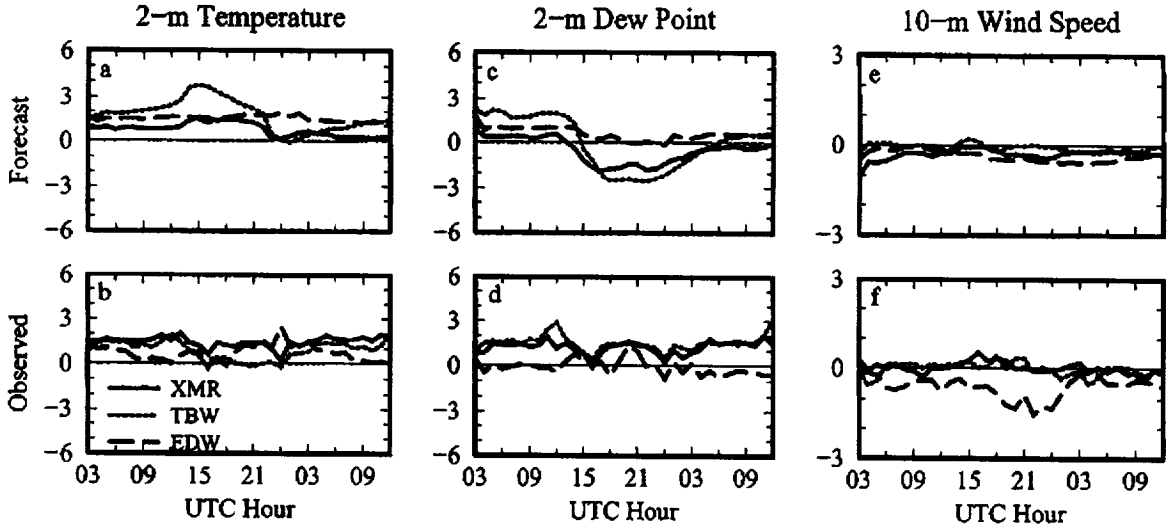

Figure 6. Same as Fig. 4 but for the 1997 cool season.

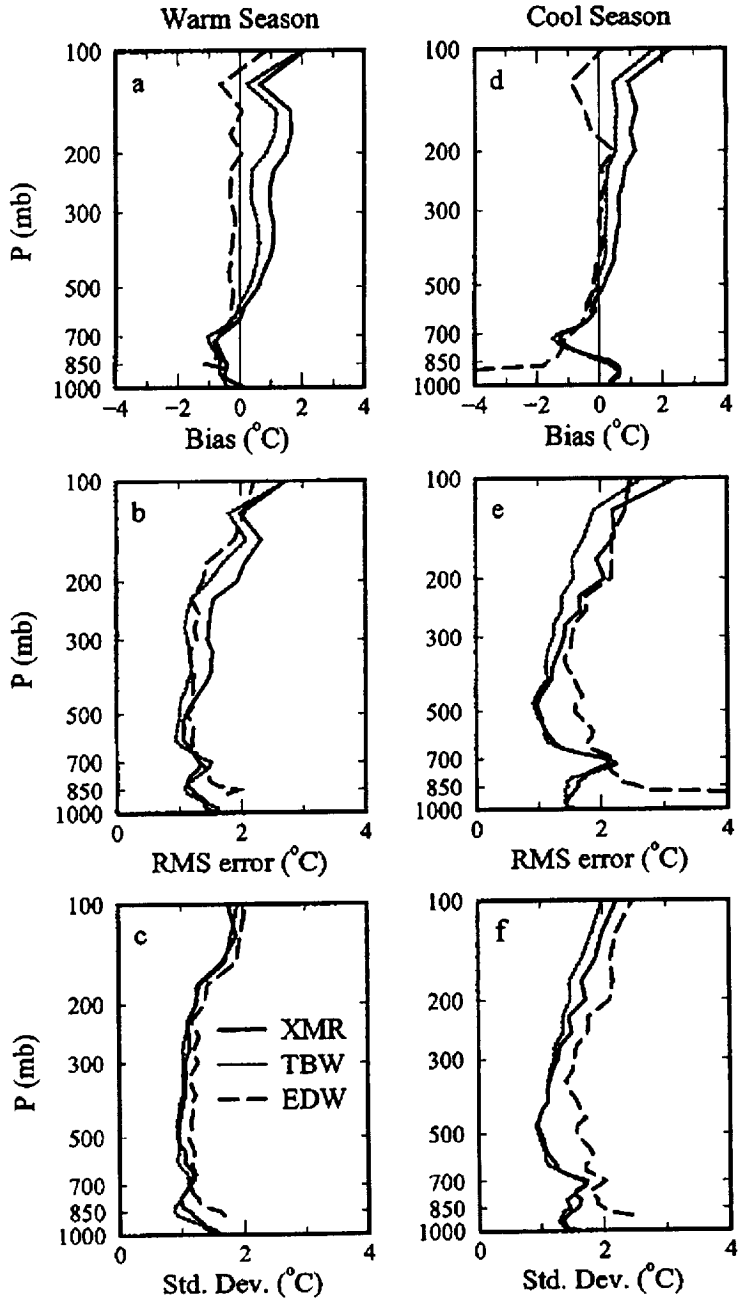

Figure 7. Bias, RMS error, and error standard deviation (°C) of meso-eta temperature forecasts plotted as a function of pressure level for XMR (solid), TBW (dotted), and EDW (dashed). Errors for the warm season are shown in the left column (panels a-c) while errors for the cool season are shown in the right column (panels d-f).

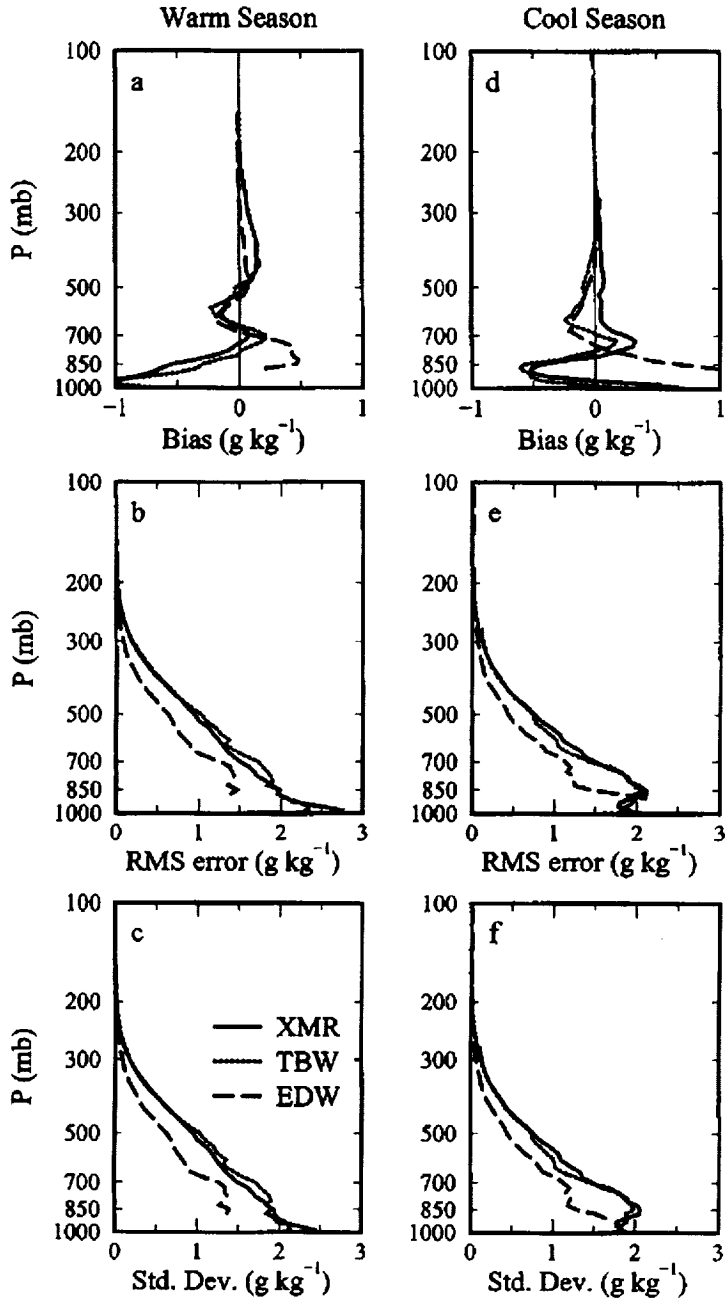

Figure 8. Same as Fig. 7 but for mixing ratio (g kg^{-1}).

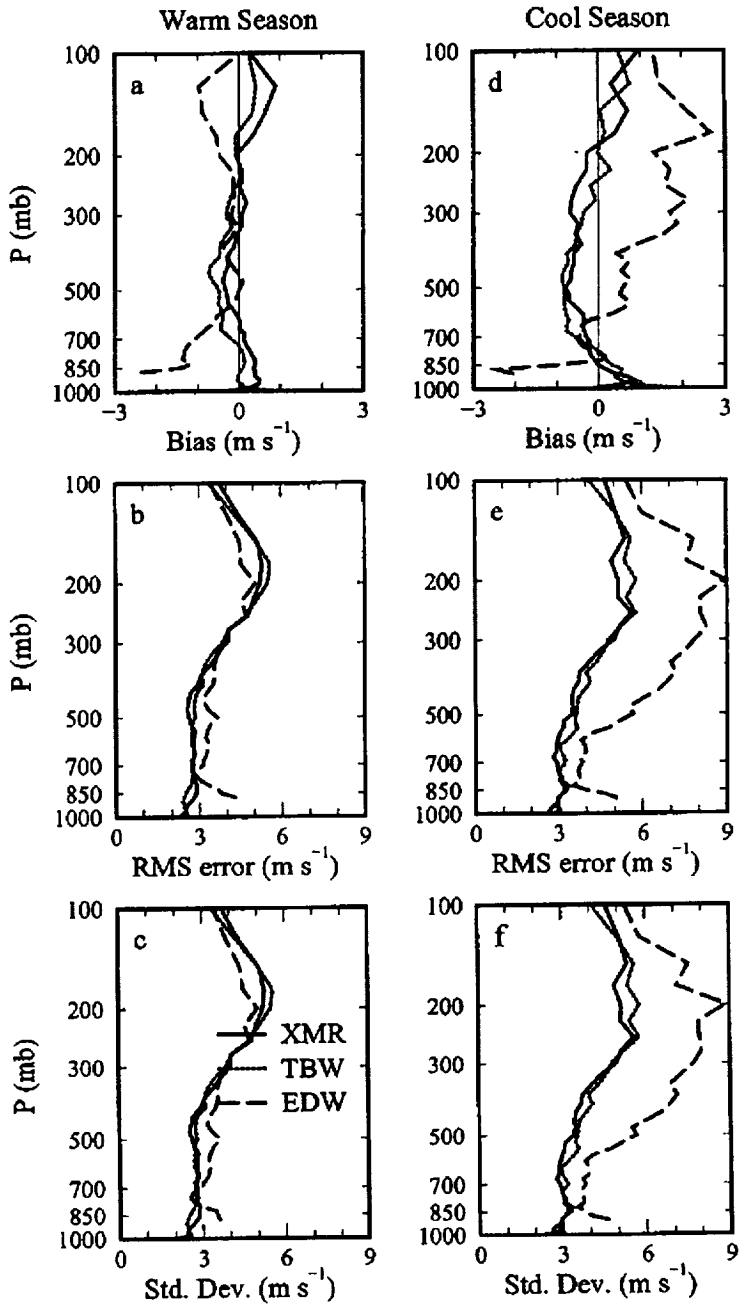

Figure 9. Same as Fig. 7 but for wind speed (m s^{-1}).

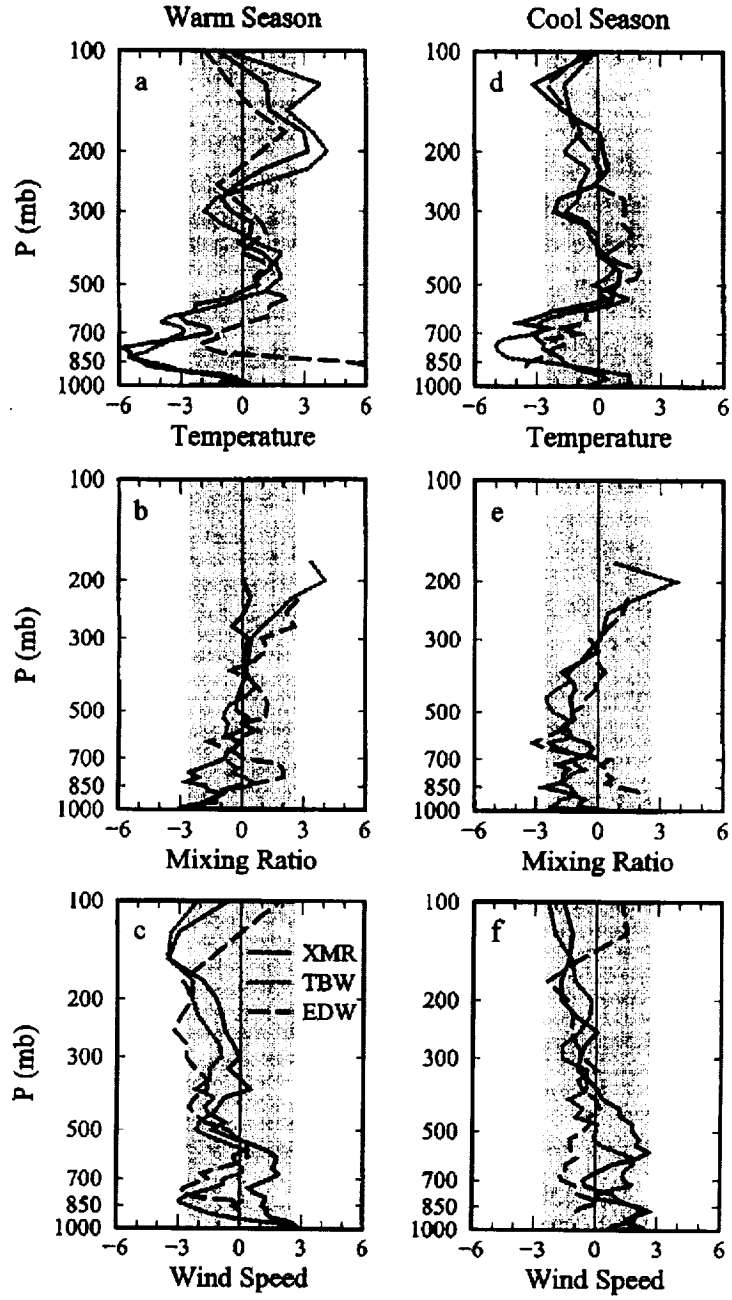

Figure 10. Paired Z statistic plotted as a function of pressure level for meso-eta forecast errors at XMR (solid), TBW (dotted), and EDW (dashed). The nondimensional statistic is shown for temperature (a, d), mixing ratio (b, e) and wind speed (c, f). Warm season values are shown in the left column (panels a-c) while cool season values are shown in the right column (panels d-f). Paired Z values that lie outside the shaded region indicate that 24-h systematic error growth is statistically significant at the 99% confidence level (see Appendix).